Estimación del nivel de habilidad en sistemas tutores inteligentes utilizando una metodología multiatributo

Sonia G. Sosa-León[1,2], Julio Waissman[2], Jose A. Olivas[1] y Manuel E. Prieto[1]

[1] Departamento de Tecnologías y Sistemas de Información. Universidad de Castilla-La Mancha. Ciudad Real, España
[2] Departamento de Matemáticas. Universidad de Sonora. Hermosillo, México
sonia@mat.uson.mx

Resumen. Para el funcionamiento ideal de un sistema tutor inteligente es indispensable poder estimar el nivel de habilidad de los estudiantes de acuerdo a objetivos complejos de aprendizaje. En este trabajo se propone una arquitectura para la evaluación del nivel de habilidad del estudiante, basada en la teoría de la utilidad multiatributo, utilizando como operador de agregación a la integral de Choquet. El método toma en cuenta los objetivos de aprendizaje planteados por el tomador de decisiones (académicos, representantes instituciones, etc.) representados por relaciones complejas que se pueden dar entre los criterios considerados para la evaluación.

Palabra Clave: Sistemas tutores inteligentes, MAUT, Integral de Choquet.

1 Introducción

Los sistemas tutores inteligentes (ITS por sus siglas en inglés) dan un complemento a la educación formal, permitiendo a los estudiantes reforzar sus conocimientos a su ritmo, con libertad de horarios y de lugar de trabajo. La personalización de las actividades propuestas por el ITS es un factor clave para su aprovechamiento [1]. En [2] se señalan tres problemas principales en la personalización de actividades: la estimación de la dificultad de los ejercicios, el algoritmo de personalización y la estimación del nivel de habilidad del estudiante. En este trabajo se propone un método para la estimación del nivel de habilidad del estudiante, para su aplicación en la plataforma Sofía XT.

Sofía XT es una plataforma desarrollada e implementada en México como complemento a la enseñanza clásica, buscando mejorar las habilidades matemáticas y fomentar el interés en ellas para los grados de enseñanza básica. Actualmente se utiliza en diferentes instituciones, tanto públicas como privadas [3]. El desempeño de Sofía XT depende en gran medida de la medición de la evolución del estudiante, por lo que la estimación del nivel de habilidad del estudiante es fundamental.

Establecer los criterios que permitan evaluar el nivel de habilidad de los estudiantes es una tarea compleja que requiere el manejo de niveles de abstracción elevados [4]. Típicamente este cálculo se realiza utilizando el modelo de Rasch [5] o una versión modificada de éste [6], sin embargo estos métodos no se ajustan a los criterios particulares de cada institución. Por ejemplo, en educación matemática hay estudios que muestran que la ansiedad matemática puede causar bajos desempeños [7] mientras que otros sustentan un punto de vista opuesto en el cual, es el bajo desempeño matemático lo que causa la ansiedad matemática [8].

La teoría de la utilidad multiatributo (MAUT) es una metodología adecuada para asignar un valor global a una situación que puede ser observada desde diferentes puntos de vista, tomando en cuenta, además, que dichos puntos de vista pueden ser de muy diferente naturaleza [9].

La contribución principal de este trabajo es la propuesta de una arquitectura para la evaluación del nivel de habilidad de los estudiantes después de realizar actividades en Sofia XT, en el cual los tomadores de decisión (académico, responsables institucionales o escolares, …) puedan expresar las relaciones complejas por las cuales medir el nivel de habilidad de acuerdo a diferentes criterios.

El artículo se organiza de la siguiente manera: En la siguiente sección se presenta la idea general de la metodología MAUT de ayuda a la decisión. En la sección 3 se presenta la operación general de la plataforma y la arquitectura del sistema de evaluación del nivel de habilidad propuesto. En las secciones 4 y 5 se presentan los operadores propuestos para el cálculo de los indicadores de desempeño marginales y globales respectivamente. Por último, se presentan las conclusiones, así como los trabajos futuros.

2 Metodología MAUT

Sea $X \subseteq X_1 \times X_2 \times \cdots \times X_n$, $n \geq 2$, un conjunto de situaciones de interés descrito por un conjunto $N = \{1, 2, \dots, n\}$ de atributos, los cuales se denominan índices. La intención, dentro de la MAUT, es modelar las preferencias de un tomador de decisión (DM) representada por un operador binario $\succcurlyeq$ en X, por medio de una función global de utilidad $U: X \rightarrow [0,1]$ tal que

$$x \succcurlyeq y \Leftrightarrow U(x) \geq U(y), \quad \forall x, y \in X.$$

La función U es generalmente determinada a partir de la información proveniente del DM mediante la cual expresa sus preferencias sobre un subconjunto relativamente pequeño y prototípico de situaciones y de atributos. La función global de utilidad U es entonces una representación numérica de la relación $\succcurlyeq$ en X y puede utilizarse como un modelo del conocimiento experto del DM.

El operador binario de preferencia $\succcurlyeq$ es una relación completa y transitiva. El modelo MAUT más general es el llamado modelo transitivo decomposable, propuesto por Krantz [7]. En este, U se obtiene por

$$U(x) = g(u_1(x_1), u_2(x_2), \dots, u_n(x_n)), \qquad \forall x = (x_1, x_2, \dots, x_n) \in X,$$

donde las funciones $u_i: X_i \rightarrow [0,1]$ se conocen como funciones marginales de utilidad, y la función no decreciente $g: [0,1]^n \rightarrow [0,1]$ es llamada operador de agregación.

La capacidad de descomponer el proceso de calificación en dos etapas depende de la selección de las funciones u_i las cuales deben de ser conmensurables. Dos funciones de utilidad marginal son conmensurables si los criterios son similares en cuanto a magnitudes, esto es que, de acuerdo al tomador de decisiones, los atributos i y j cumplen en el mismo grado con sus criterios marginales.

3 Arquitectura propuesta

La plataforma Sofía XT funciona en términos generales de la siguiente manera:

1. En base a la información de registro, al estudiante se le asigna un nivel de habilidad por cada concepto y en base a su nivel, el sistema habilita los conceptos que el estudiante puede seleccionar.
2. Una vez que el estudiante selecciona un concepto, se genera una actividad, que consiste en 5 ejercicios en orden creciente de dificultad, a partir del conjunto de ejercicios asociados al concepto.
3. El estudiante realiza la actividad y se reevalúa el nivel de habilidad del estudiante sobre el concepto y se asignan recompensas para el estudiante, en relación a su progreso.
   a. Si el estudiante no cumple con un nivel de habilidad mínimo, se le presenta la opción de seleccionar uno de los conceptos antecedentes al actual.
   b. En caso que el estudiante tenga un nivel de habilidad alto en este concepto, se le presenta una selección de conceptos adecuados a su nivel.
   c. En otro caso se mantiene la misma selección de conceptos.

Un método de estímulos y recompensas busca guiar al estudiante, al mismo tiempo que le brinda la libertad de tomar decisiones sobre su aprendizaje, convirtiéndolo en un sujeto activo en la toma de decisión. Este método de estímulos funciona basado en un avatar, y la otorgación de créditos para que los estudiantes personalicen su avatar. De acuerdo a lo observado por los profesores, el sistema de créditos ha estimulado en los estudiantes el uso de la plataforma.

La base para la operación correcta de la plataforma es la estimación del nivel de habilidad del estudiante, la cual se realiza tomando en cuenta el nivel de habilidad anterior y la evaluación de la última actividad. En dicha evaluación es donde se requiere expresar los objetivos de aprendizaje.

En la Fig. 1 se presenta un esquema con la adaptación del método MAUT, propuesto para la calificación de las actividades resueltas por estudiantes. Una vez que un estudiante resuelve una actividad, se genera un vector de cinco entradas $X_A = [x_1, x_2, x_3, x_4, x_5,]$, las cuales a su vez son estructuras complejas de datos. La estructura de la entrada incluye como información, el identificador del estudiante que la realizó $e(x_k)$, la dificultad del problema $d(x_k)$, el tiempo en segundos que tardó en realizar el problema $t(x_k)$ y los intentos que requirió el estudiante para realizar el problema $i(x_k)$, además de la información genérica de la actividad. Esta información se procesa para obtener dos conjuntos de datos: por una parte, se obtiene un conjunto de atributos específicos a la actividad realizada, los cuales servirán para la calificación como atributos en un proceso MAUT; por otra, un conjunto complementario de datos (concepto, grado, área de conocimiento), para establecer los criterios de los indicadores de desempeño marginal. Los atributos específicos a la actividad son la dificultad, el tiempo utilizado y el número de intentos realizados por cada uno de los ejercicios, definidos por $\bar{X}_k = (\mathrm{d}(x_1), t(x_1), i(x_1), \dots, \mathrm{d}(x_5), t(x_5), i(x_5))$.

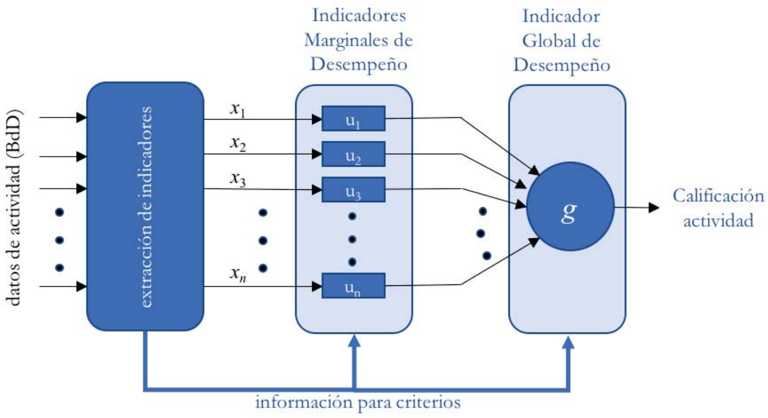

Fig. 1. Arquitectura tipo MAUT de evaluación de actividades

Estos atributos son llevados a las diferentes funciones u_i las cuales se designan como indicadores marginales de desempeño (MPI). Los MPI se pueden tratar y analizar como funciones de pertenencias, las cuales definen 15 subconjuntos borrosos. Si bien, se calcula cada uno independiente de las otras mediciones de la actividad, debido al requerimiento de conmesurabilidad de índices en el proceso MAUT, todos los MPI similares serán calculados de la misma forma. Así, los MPI se calculan como funciones de pertenencia difusas parametrizadas a partir del conjunto complementario de datos.

Una vez que los datos se han tratado, se obtiene un vector de indicadores, los cuales se van a utilizar para generar un indicador global de desempeño. Los operadores de agregación g son funciones globalmente estrictamente crecientes, tales que $g(0, \dots, 0) = 0$ y $g(1, \dots, 1) = 1$. En las secciones siguientes se desarrollan los indicadores marginales y global de desempeño propuestos.

4 Indicadores marginales de desempeño

Se desarrollaron 3 tipos básicos de funciones de pertenencia borrosas: la dificultad, el tiempo utilizado para contestar el reactivo y el número de intentos realizado.

El grado de dificultad de cada ejercicio se calcula en forma estadística, modelando el grado de dificultad de los reactivos como una variable aleatoria normal $\mathcal{N}(0,1)$, donde $d(x_k) = 0$ es la dificultad media de los ejercicios. A partir de esta información, el MPI respecto a la dificultad de un reactivo se puede realizar como una aproximación de la función de probabilidad dada por

$$u_{3k-2}(d(x_k)) = \frac{1}{2}\left(1 + \mathrm{sign}(d(x_k))\left(1 - e^{-2d(x_k)/\pi}\right)^{1/2}\right).$$

Para el cálculo de los MPI correspondientes al tiempo que le tomó al estudiante realizar el reactivo, es necesario parametrizar en relación a los datos complementarios (el concepto, subtema y grado), debido a que estos difieren entre sí. Como podemos observar en el ejemplo de la Fig. 2, la distribución de los tiempos presenta una regularidad entre todos los conceptos, variando su valor medio en función del nivel de dificultad del concepto.

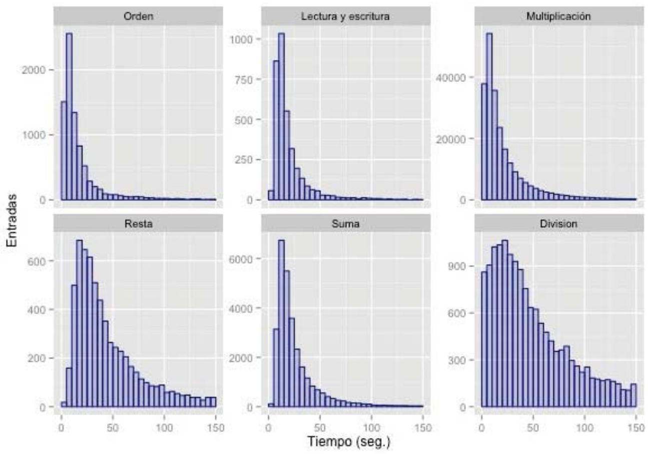

Fig. 2. Distribución del tiempo de resolución del subtema *Números naturales de cuatro cifras*

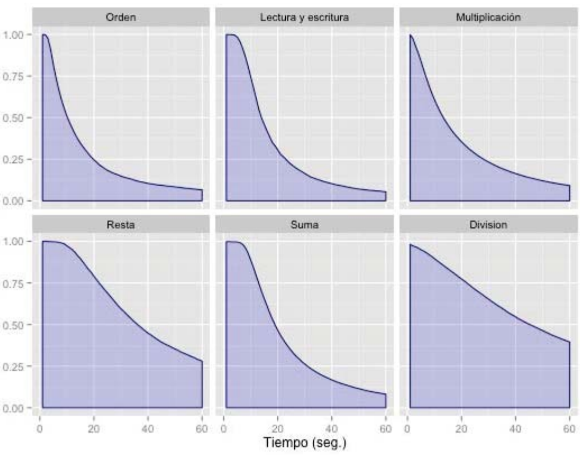

Fig. 3. Índices marginales de desempeño en tiempo para el subtema *Números naturales de cuatro cifras*

Sea φ_t^C el número de ocurrencias en que el conjunto de entradas históricas pertenecientes al concepto C tuvo un intervalo de tiempo de resolución t, en segundos. El cálculo de los MPI de tiempo se realiza mediante

$$u_{3k-1}\left(t\left(x_{i,k}\right)\right) = 1 - \frac{\sum_{k=1}^{\lfloor t(x_k) \rfloor} \varphi_k^C}{\sum_{\forall t} \varphi_t^C}.$$

Si bien el índice propuesto solo contempla tiempos en una granularidad de segundos, esta parece ser una escala de detalle suficientemente suave, tal como se aprecia en la Fig. 3.

En la Fig. 4 podemos ver el número de intentos de los 6 conceptos que conforman el subtema *Números naturales de cuatro cifras* de tercer grado. La gran mayoría de las entradas se resolvió en el primer intento como se puede observar. Otro comportamiento interesante de observar es el número de entradas que registran 0 intentos. En la plataforma Sofía XT los estudiantes pueden decidir saltar un ejercicio sin resolverlo, a pesar de las ayudas ofrecidas. Cada vez que un estudiante salta un ejercicio, su número de intentos se marca como 0. Si bien en algunos conceptos revisados existe un número mayor que 4 de intentos significativos, estos nunca (en todos los conceptos de la plataforma) son mayores a 10.

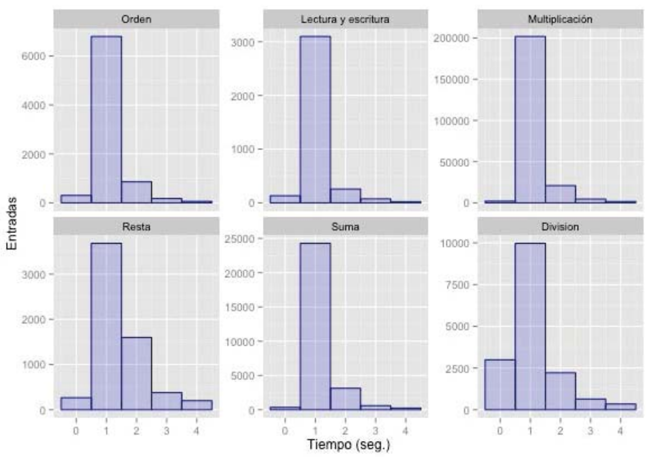

Fig. 4. Distribución de intentos del subtema *Números naturales de cuatro cifras*

Sea θ_i^C el número de entradas del concepto C que se registró con i intentos, entonces el índice de desempeño marginal se define como

$$u_{3k}(i(x_k)) = \begin{cases} i(x_k), & \text{si } i(x_k) < 2, \\ \left(\dfrac{\theta_{i(x_k)}^C}{\sum_{l=0}^{10} \theta_l^C}\right)^{1/r}, & \text{en otro caso,} \end{cases}$$

donde $r \geq 1$ es una constante de autoestima. Entre mayor sea el valor de r, menor será la exigencia del entorno de aprendizaje para que los estudiantes resuelvan los ejercicios al primer intento. Como se observa en la función, solamente si se resuelve el ejercicio en el primer intento, el valor del índice será de 1 y solamente si el estudiante decidió no resolver el ejercicio se asignará el valor de 0.

5 Indicador global de desempeño

Una vez procesada la información de la actividad por los MPI se obtienen quince valores de desempeño marginal. Éstos se arreglan de la forma siguiente:

$$u = (d_1, t_1, i_1, \dots, d_5, t_5, i_5),$$

donde d_k, t_k y i_k representan al valor del MPI de dificultad, tiempo e intentos de la entrada k de la actividad a calificar. Usualmente, las actividades de estudiantes se evalúan por medias ponderadas. Esto implica que las relaciones entre la dificultad, el tiempo y los intentos de un mismo ejercicio, o entre los ejercicios de la actividad se consideran irrelevantes para la evaluación. Sin embargo, en [10] y [11] se muestra con ejemplos que aún en casos muy sencillos, esta hipótesis no se cumple generalmente.

En este trabajo se propone como operador de agragación la integral de Choquet. Sea $N = \{1, 2, \dots, n\}$ el conjunto de n índices, de manera que cada índice represente a un atributo de una situación $x = (x_1, x_2, \dots, x_n) \in X$. Una capacidad borrosa es una función del conjunto potencia de N al intervalo unitario, $\mu: \mathcal{P}(N) \rightarrow [0, 1]$ que satisface las condiciones siguientes:

- $\mu(\emptyset) = 0, \mu(N) = 1$ (condiciones de frontera);
- para todo $S, T \subseteq N$, tal que $S \subseteq T$ entonces $\mu(S) \leq \mu(T)$ (monotonía).

Para cada subconjunto de índices $S \subseteq N$, el número $\mu(S)$ representa el peso o importancia de S. La monotonía de μ significa que el peso o importancia de un subconjunto de índices no puede disminuir si se agregan nuevos índices. La integral de Choquet respecto a una capacidad μ es una función $C_\mu: [0, 1]^n \rightarrow [0, 1]$ definida por

$$C_\mu(x) \coloneqq \sum_{i=1}^{n} x_{\sigma(i)} \left[\mu\left(A_{\sigma(i)}\right) - \mu\left(A_{\sigma(i+1)}\right)\right],$$

donde $\sigma = (\sigma(1), \dots, \sigma(n))$ es una permutación de N tal que $x_{\sigma(1)} \leq x_{\sigma(1)} \leq \cdots \leq x_{\sigma(n)}$. Igualmente, para todo $i \in N$, $A_{\sigma(i)} = \{\sigma(i), \dots, \sigma(n)\}$ y $A_{\sigma(n+1)} = 0$.

El análisis del comportamiento de la integral de Choquet en cuanto a relevancia de criterios e interacciones, se hace mediante dos de los índices más reportados en la literatura. Para evaluar la importancia de cada uno de los atributos respecto a los demás, utilizamos los valores de Shapley, definidos como

$$\phi_\mu(i) \coloneqq \sum_{T \subseteq N - \{i\}} \frac{(n - |T| - 1)!\,|T|!}{n!} \left(\mu(T \cup \{i\}) - \mu(T)\right),$$

y pueden verse como el valor promedio de la contribución marginal de i a todas las capacidades.

La interacción básica entre dos índices i y j se puede obtener calculando la diferencia $[\mu(\{i,j\})]$ - $[\mu(\{i\}) + \mu(\{j\})]$. Si la diferencia es positiva, se dice que son complementarios; si la diferencia es 0 se dice que son independientes; y si la diferencia es negativa, se dicen redundantes. La idea básica del valor de Shapley se puede generalizar a la relación con los dos índices, permitiendo calcular un valor de interacción de la forma

$$I_\mu(i,j) = \sum_{T \subseteq N-\{i,j\}} \frac{(n-|T|-2)!\,|T|!}{(n-1)!} [\mu(T \cup \{i,j\} - \mu(T \cup \{i\}) - \mu(T \cup \{j\}) + \mu(T)] \, .$$

El valor $I_\mu(i,j) \in [-1,1]$ puede por lo tanto ser considerado como el promedio de la interacción marginal de i y j, cuyo valor es 1 si existe una complementariedad máxima y -1 si los índices son completamente redundantes.

Frente a problemas de evaluación, es poco factible solicitar a la institución o pedagogos reglas generales que permitan obtener el operador de agregación. Para los expertos puede resultar más fácil expresar su conocimiento en casos concretos. En este trabajo se propone un método interactivo donde los expertos expresen su conocimiento como relaciones de preferencias, que irán tomando una complejidad creciente a medida que se interacciona con el método de adaptación. El experto expresa su conocimiento a partir de ejemplos de actividades que pueden ser reales o casos prototípicos.

Sea u y u' dos actividades prototípicas descritas por un vector de MPI de dimensión 15, se dice que u debe estar mejor calificada u' y se denota $u \succcurlyeq u'$ si y solo si

$$C_\mu(u) - C_\mu(u') \geq \delta_c,$$

donde δ_C es una constante pequeña. De la misma forma, se dice que dos actividades deben tener una calificación parecida (denotada por $u^{(p)} \sim_A u^{(q)}$) siempre que $\left|C_\mu(u) - C_\mu(u')\right| \leq \delta_c$.

Una vez establecido el conjunto de casos particulares y un orden parcial de ellos, se establece un orden entre los valores de importancia de los índices, lo que permite expresar relaciones de importancia entre la dificultad de los ejercicios y la importancia del tiempo respecto a los intentos en cada ejercicio.

Se dice que un índice i se prefiere respecto a un índice j ($i \succcurlyeq_N j$) si $\phi_\mu(i) - \phi_\mu(j) \geq \delta_{Sh}$ donde δ_{Sh}es una constante muy pequeña. De la misma forma dos índices i y j tienen una importancia semejante ($i \sim_N j$) si $|\phi_\mu(i) - \phi_\mu(j)| \leq \delta_{Sh}$.

Si es necesario, se pueden establecer criterios que fijen las interdependencias entre los diferentes criterios. La relación entre los índices i y j se obtiene directamente al imponer el criterio $I_\mu(i,j) < 0$ para redundancia $I_\mu > 0$ para complementabilidad y $|I_\mu(i,j)| < \epsilon$ para independencia. Por otra parte, las relaciones de preferencia entre relaciones de índices, tales como $i,j \succcurlyeq_I k,l$ están dadas por $I_\mu(i,j) - I_\mu(k,l) \geq \delta_I$.

El proceso de especificación se hace en forma incremental, buscando representar cada vez un objetivo de aprendizaje más complejo. El método de construcción de g se desarrolló en el lenguaje estadístico R usando librería *kappalab* [9].

Con el fin de ilustrar el proceso de generación de operador de agregación se presenta un ejemplo ilustrativo, el cual se basa en cinco casos prototípicos representados en el cuadro 1. El orden de preferencia establecido por el tomador de decisiones es $\boldsymbol{e}_1 \succcurlyeq \boldsymbol{e}_2 \succcurlyeq \boldsymbol{e}_3 \succcurlyeq \boldsymbol{e}_4 \succcurlyeq \boldsymbol{e}_5$. Este criterio busca fomentar la competitividad matemática.

Cuadro 1. Casos prototípicos del ejemplo ilustrativo

Estudiante	Índices marginales de desempeño														
	d_1	t_1	i_1	d_2	t_2	i_2	d_3	t_3	i_3	d_4	t_4	i_4	d_5	t_5	i_5
$\boldsymbol{e_1}$	0.2	0.4	0.9	0.4	0.4	0.9	0.6	0.4	0.5	0.8	0.4	0.5	1.0	0.4	0.5
$\boldsymbol{e_2}$	0.2	0.8	0.9	0.4	0.8	0.5	0.6	0.4	0.9	0.8	0.4	0.5	1.0	0.4	0.5
$\boldsymbol{e_3}$	0.2	0.8	0.5	0.4	0.8	0.9	0.6	0.4	0.9	0.8	0.4	0.5	1.0	0.4	0.5
$\boldsymbol{e_4}$	0.2	0.8	0.5	0.4	0.8	0.5	0.6	0.4	0.9	0.8	0.4	0.9	1.0	0.4	0.5
$\boldsymbol{e_5}$	0.2	0.8	0.5	0.4	0.4	0.5	0.6	0.8	0.5	0.8	0.8	0.9	1.0	0.4	0.9

Una vez parametrizada la integral de Choquet se evalúan los casos prototípicos obteniendo la estimación del nivel de habilidad para cada estudiante siguiente: $g(\boldsymbol{e}_1) = 0.9$, $g(\boldsymbol{e}_2) = 0.8$, $g(\boldsymbol{e}_3) = 0.7$, $g(\boldsymbol{e}_4) = 0.6$ y $g(\boldsymbol{e}_5) = 0.5$ respectivamente.

Con el fin de entender el comportamiento de la evaluación, se revisan los valores de Shapley en cada criterio y los índices de interacción. Dado que se desea establecer un criterio para que la plataforma promueva la competitividad matemática, se revisan principalmente los índices que se relacionan con el número de intentos. Al obtener los valores de Shapley se encuentra que $\phi_\mu(i_1) = 0.375$, $\phi_\mu(i_2) = \phi_\mu(i_3) = 0.125$ y los valores de Shapley para los problemas con mayor dificultad $\phi_\mu(i_4) = \phi_\mu(i_5) = 0$.

Esta conducta es inaceptable desde el punto de vista de la competitividad matemática, ya que no le da importancia a requerir muchos intentos en los ejercicios de mayor dificultad. Con este fin se agrega al algoritmo de parametrización del operador de agregación la consideración

$$i_1 \sim_N i_2 \sim_N i_3 \sim_N i_4 \sim_N i_5.$$

Una vez obtenida la nueva parametrización, se obtienen valores de Shapley $\phi_\mu(i_k) = 0.125$ para $k = 1, \dots, 5$. Al recalcular los niveles estimados de habilidad obtenemos lo siguiente: $g(\boldsymbol{e}_1) = 0.68$, $g(\boldsymbol{e}_2) = 0.63$, $g(\boldsymbol{e}_3) = 0.58$, $g(\boldsymbol{e}_4) = 0.53$ y $g(\boldsymbol{e}_5) = 0.48$ respectivamente. Estas evaluaciones cumplen con todos los criterios establecidos y se ajustan mejor a las evaluaciones que consideran correctas los tomadores de decisiones. Es interesante resaltar que, al estudiar la relación de interacción entre índices, se encontró que el número de intentos entre los diferentes problemas son independientes entre sí salvo por los ejercicios 3 y 5 en el cual $I_\mu(i_3, i_5) = 0.25$.

6 Conclusiones y trabajos futuros

En este trabajo se propone un método basado en la metodología MAUT para la evaluación del nivel de habilidad de un estudiante, en el cual los objetivos de aprendizaje se puedan establecer en términos de comparación de casos reales o prototípicos. El método propuesto utiliza la información histórica de la plataforma para establecer los índices de desempeño marginal, así como el conocimiento experto y las preferencias de los tomadores de decisiones, a través de un método interactivo para construir un operador de agregación. Para esto último, se adaptó la metodología MAUT que permite expresar un conocimiento experto en función de relaciones particulares tanto entre niveles de habilidad de estudiantes como entre los criterios considerados para la evaluación.

Actualmente el método propuesto se encuentra en etapas de prueba en la plataforma Sofia XT. Con esta propuesta se espera poder personalizar la plataforma de enseñanza a las necesidades y objetivos de cada institución de enseñanza que la utilice.

Agradecimientos

Este trabajo ha sido parcialmente financiado por la Agencia Estatal de Investigación (AEI) y el Fondo Europeo de Desarrollo Regional (FEDER) dentro del proyecto MERINET: TIN2016-76843-C4-2-R (AEI/FEDER, UE). Sonia G. Sosa-León, agradece el apoyo otorgado por PRODEP (México) mediante beca con folio UNISON-344. Julio Waissman agradece a la Universidad Castilla – La Mancha por su financiamiento a través del programa de apoyo a profesores visitantes.

Referencias

1. Wauters, K., Desmet, P., Van Den Noortgate, W.: Adaptive item-based learning environments based on the item response theory: Possibilities and challenges. Journal of Computer Assisted Learning, 26(6), 549–562 (2010).
2. Klinkenberg, S., Straatemeier, M., Van Der Maas, H. L.: Computer adaptive practice of Maths ability using a new item response model for on the fly ability and difficulty estimation. Computers and Education, 57(2), 1813–1824 (2011).
3. Cacho-Carranza, Y.: Sofía XT, una web para mejorar el aprendizaje matemático en niños. *Agencia Informativa CONACYT AIC*. Ciudad de México. http://conacytprensa.mx/index.php/tecnologia/tic/, ultimo acceso 2017/06/01.
4. Kamvysi, K., Gotzamani, K., Andronikidis, A., Georgiou, A. C.: Capturing and prioritizing students' requirements for course design by embedding Fuzzy-AHP and linear programming in QFD. European Journal of Operational Research, 237(3), 1083–1094 (2014).
5. Rasch, G.: Probabilistic models for some intelligence and achievement tests. Danish Institute for Educational Research, Copenhagen (1960).
6. Maris, G., van der Maas, H.: Speed-Accuracy Response Models: Scoring Rules based on Response Time and Accuracy. Psychometrika, 77(4), 615–633 (2012).
7. Chinn, S.: Mathematics Anxiety in Secondary Students in England. *Dyslexia*, *15*, 61–68 (2009).
8. Ma, X., Xu, J.: The causal ordering of mathematics anxiety and mathematics achievement: A longitudinal panel analysis. Journal of Adolescence, 27(2), 165–179 (2004).
9. Bigaret, S., Meyer, P.: Evaluation and Decision Models with Multiple Criteria: Case Studies. Springer-Verlag, Berlin Heidelberg (2015).
10. Shieh, J. I., Wu, H. H., Liu, H. C.: Applying a complexity-based Choquet integral to evaluate students' performance. Expert Systems with Applications, 36(3):5100–5106, (2009).
11. Grabisch, M., Kojadinovic, I., Meyer, P.: A review of methods for capacity identification in Choquet integral based multi-attribute utility theory. Applications of the Kappalab R package. European Journal of Operational Research, 186(2):766–785, (2008).